sciendo
DOI 10.2478/jazcas-2021-0056

A CORPUS OF CZECH ESSAYS FROM THE TURN OF THE 1900s

PETR POŘÍZKA
Department of Czech Studies, Faculty of Arts, Palacký University, Olomouc, Czech Republic

POŘÍZKA, Petr: A corpus of Czech essays from the turn of the 1900s. Journal of Linguistics, 2021, Vol. 72, No 2, pp. 618 – 630.

Abstract: A literary essay is an interesting unit for language analyses, as its stylistic means often exceed the boundaries of the genre of an artistic essay. The article presents a new corpus of Czech literary essays covering approximately fifty years from 1890 to 1940. Along with the characterisation of the corpus and its annotation, the paper focuses on the TXM corpus tool: In the second part of the study, we use selected texts to conduct an analysis of seven various authors through multidimensional cluster analysis, factorial correspondence analysis and a specificity score. The main parameter of the analyses was usage of parts of speech in texts by individual authors. At present, the Corpus of Czech Essays contains 40 essayist titles written by 15 authors covering various topics (music, visual arts, theatre, literature, etc.).

Keywords: annotation, corpus, corpus linguistics, quantitative analysis, literary essay, multidimensional analysis, orthography, specificity score, TXM

1 INTRODUCTION

At present, Czech linguistics already has a number of corpora available, covering a range of areas with regard to both temporal and typological or genre characteristics. Some textual areas or language periods are, however, covered to a lesser extent or are awaiting processing. One interesting period in the development of standard Czech is the turn of the 1900s, when the views of standard Czech and its orthographic form were established. Attitudes of linguists' changed turbulently during this time. One might mention in this context various polishings representing purist efforts and tendencies, followed by the attempts at stabilization of standard Czech through grammar guidebooks and rulebooks (especially that by J. Gebauer), and finally the Prague Linguistic Circle which regarded the form of standard Czech as one of the key topics.

Czech literary essays from this time illustrate this period of development and also have an indisputable literary-aesthetic value. Although the genre is rather narrowly focused, the options for their utilization for language analyses are undoubtedly wider, since the language and stylistic means used by the authors included in the corpus often exceed the genre of the literary essay. The language of

these texts oscillates, reflecting the means of multiple functional styles: artistic, scientific, journalistic, rhetorical (and partially colloquial); perhaps the only one not involved is the administrative style. This wide range and certain typical tendency to overstep the borders and blend individual functional styles are also confirmed by the only fairly comprehensive anthology of Czech literary essays published in two volumes ([1], [2]). Opelík structures the second volume of the anthology into chapters covering program, portrait, poetological, reprimanding and reflexive essays [2]. The delimitation of the essayist style, as a relatively autonomous unit within the system of functional styles, was first attempted by Havránek in his commentary on functional differentiation of language (1932), although he did not classify it among the basic styles (cf. [3], [4]). Hausenblas [5] classified it as a *complex* style (in contrast to *simplex* styles) and within present-day theory of functional styles, it is classified as a *secondary* style, cf. [6]. When determining the style-based essence of the essay, Jedlička ([4], [7]) pointed out (a) its characteristic tendency to weakening of terminological saturation of a text and (b) a significant proportion of the register of highly formal and dynamic language means. Mistrík [8] defined the borderline character of the literary essay in relation to the (i) scientific, (ii) journalistic and (iii) artistic style. Literary essays are interesting even from the perspective of the lexical means used: formal, expressive and even exclusive means, nonce words, figurative expressions, etc. It was particularly the above-described linguistic character of literary essays – its multifaceted and borderline nature, oscillation among multiple functional styles and mutual blending of language means from various styles – that encouraged us to create a corpus of Czech literary essays (hereinafter also CCE).

2 CHARACTERISTICS OF THE CORPUS

A corpus of this kind must necessarily include texts written by the founder of Czech literary essays F. X. Šalda, the "poet" of Czech essays Otokar Březina, as well as philosophical essays by Ladislav Klíma. The corpus incorporates almost 6 thousand pages of various types of texts (fictional, scientific, journalistic) from various areas (music, visual arts, theatre, literature, etc.). In total, the corpus presently contains 40 books of essays by 15 authors (i.e. on average two to three books for every author) published between 1890–1937. The following authors are included in the present version of CCE: Otokar Březina, Josef Čapek, Karel Čapek, Jaroslav Durych, Otakar Hostinský, Jiří Karásek, Ladislav Klíma, F. V. Krejčí, Jiří Mahen, Miloš Marten, Vilém Mathesius, Arne Novák, Arnošt Procházka, H. G. Schauer and F. X. Šalda.

2.1 Data sources and data processing

The texts included in the corpus come from several sources. The most important one is *Digital Library Kramerius* – a database of the National Library [9]. In

addition, we also used *Digital Library* of the Moravian Library [10], complemented with library loans and OCR conversion of texts into an electronic version. Along with the selection of a particular author and text, the key parameters also included a free license with respect to copyright – expired copyright protection (70+ years from the author's death) – and the version of a particular text: we used the first edition.

For the processing of data, we used the help of students within specialized seminars: each student processed a part of a particular book (ca. 100 pages). Source texts were available in two versions: (1) a set of scanned images (jpg) – the original of the book, and (2) a folder with texts after an automatic OCR conversion (txt). There was a need to make a detailed and precise manual correction for every book according to the original text, as the electronic version (ad 2) is available in *Kramerius* and *Digital Library* databases in a non-revised version – i.e. including all the errors resulting from the automatic conversion. The main editing principle was fidelity to the original. When needed, the text was supplemented with a corrector's note describing a particular change to the text. There was a need, for example, to edit words written in "spaced characters" (a common typographic practice of the particular period), i.e. for the word "u m ě n í" [a r t] (and similar cases elsewhere) it was necessary to delete the spaces between the individual characters and write the expression as "umění" [art]. The whitespace is one of the segmentation characters in corpus databases and without this editing change, the corpus manager would not process these cases as a single lexical unit, but as a sequence of five individual characters "u", "m", "ě", "n", "í". Similarly, there was a need to delete word division of the typographic layout of the book and pagination or add missing signs (typesetting mistakes), for instance "p dstata" [e sence] was corrected to "podstata" [essence] (with an inserted note indicating the missing "o" in the original).

In addition, a list of so-called *anomaly expressions* was purposefully created for every text with regard to the differences between the present-day and contemporary versions of orthography as well as due to the need for linguistic annotation of the texts – for subsequent corrections of automatic annotation (lemmatization and tagging).[1] The usage of this dictionary is wider, however, it allows for insight into the contemporary specific lexicon or the unique lexicon of a particular author (words such as *srostlivost* [a tendency to coalesce], *zvášnivělý* [impassioned], etc.) and may serve as instrument for analyses of texts from the database. The most common 'anomalies' were related to the following phenomena:

- the quantity of vocalic letters: *system* ['system', in Czech correctly "systém"], *primarní* ['primary', in Czech correctly "primární"]

[1] The accuracy or error rate of the annotation depends, among other things, on the tool dictionary. Our comprehensive list of anomaly words that are not part of these annotation dictionaries, may therefore be purposefully used for correction of errors of the automatic text annotation.

- orthographic rules for words of foreign origin, especially Latin and Greek:
  - vocalic digraphs (*aether*) [‘ether’]
  - ending *-ism* (*heroism*) [‘heroism’]
  - double consonant letters *ll* (*illuse*) [‘illusion’], *tt* (*marionetty*) [‘marionettes’], *rr* (*korrelata*) [‘correlates’ – plural noun], *ss* (*associace*) [‘association’], *mm* (*summa*) [‘sum’], *ff* (*affirmující*) [‘affirmating’ – present participle], *kk* (*akkumulace*) [‘accummulation]
  - other phenomena: *th* (*hypothesa*) [‘hypothesis’], *s/z* (*kausalita*) [‘causality’], *ks* instead of *x* (*ekstase*) [‘ecstasis’], *k* instead of *ch* (*karakter*) [‘character’], *qu* instead of *kv* (*quanta*) [‘quantities’].

2.2 Tool for data mining

The main corpus manager for data mining is TXM (abbrev. Textometrie) [11]. This open-source tool was chosen for a number of reasons, for instance the following:

- Unicode – XML & TEI compatible platform
- helps to build various corpus configurations; provides a large spectrum of input formats and rich data models[2]
- has broad and complex options for qualitative-quantitative data mining
- based on the efficient CQP full text search engine and its powerful CQL query language
- has enhanced functions uncommon in other corpus managers[3]:
  - the R statistical environment [12]; provides quantitative analysis, based on R packages (including the option for additional installation of any extension package), e.g.:
    - factorial correspondence analysis
    - hierarchical cluster analysis
    - specific patterns analysis (specificities)
  - includes TIGERSearch query tool for syntactic data mining
  - applies various NLP tools on the fly on texts before analysis (e.g. TreeTagger for lemmatization and POS tagging)
  - provides scripting facilities for repetitive or lengthy tasks automation or for platform extension.

2.3 Corpus format and annotation

CCE was annotated using the open-source tool *MorphoDiTa* [13] which uses a freely accessible Czech morphological dictionary *MorfFlexCZ*[4]. The texts were

[2] For more information, see the documentation of the tool: http://textometrie.ens-lyon.fr/spip.php?rubrique64.

[3] We mean here standard non-commercial corpus managers such as NoSketch Engine, KonText, Poliqarp, etc.

[4] Available at: https://ufal.mff.cuni.cz/morfflex.

lemmatized, morphologically tagged (the Czech 15-position tagset is used – see f.n. 5) and processed into XML format. The corpus annotation is represented in XML through its elements and attributes: directly with *elements* (as a structure *s-attribute*) and/or with *attributes* of these elements (positional *p-attribute*). The basic XML format would therefore look as follows (the root element `text` contains additional metadata – author, title and year of publication of the text)[5]:

```
<?xml version="1.0" encoding="UTF-8"?>
<text author="Durych" title="Essaye" year="1931">
<s>
<w lemma="oheň" pos="N">Oheň</w>
<w lemma="lidstvo" pos="N">lidstva</w>
<w lemma="svítit" pos="V">svítí</w>
<w lemma="dva" pos="C">dvěma</w>
<w lemma="plamen" pos="N">plameny</w>
...
</s>
...
</text>
```

Explanatory note: elements `text` = root element; `s` = sentence; `w` = word; attributes `lemma` and `pos` = part-of-speech.

The second most used corpus format is WPL (word per line) based on column annotation. And if there is a need, the XML format can be converted to vertical WPL-format and imported into the TXM tool using the function *Import CQP* or imported into other corpus managers based on the Manatee system (like (No) SketchEngine and more).

This annotated format was further adjusted: we extracted some of the nominal and verbal sub-categories in order to subsequently use them for corpus analysis. Specifically, a complex tag (`tag`) was used to create separate attributes for the part of speech (`pos`), gender (`g`), number (`n`), case (`c`), person (`p`), and tense (`m`).

Cf. examples below – (1) original annotation from the tool *MorphoDiTa*, and (2) the final form of annotation following adjustments (a sample of Březina's essay *Tajemné v umění* [Mystery in Art]):

(1)

```
<s>
 <w lemma="odpověď" tag="NNFP1-----A----">Odpovědi</w>
 <w lemma="být" tag="VB-P---3P-AA---">jsou</w>
 <w lemma="věčný" tag="AAFP1----1A----">věčné</w>
...
</s>
```

[5] Within the attribute `pos` the individual parts of speech are already referred to with their usual abbreviations of the Czech tagset. For more information, see: https://ufal.mff.cuni.cz/pdt/Morphology_and_Tagging/Doc/hmptagqr.html.

(2)

```
<s>
 <w lemma="odpověď" pos="N" tag="NNFP1-----A----" g="F" n="P"
  c="1" p="-" m="-">Odpovědi</w>
 <w lemma="být" pos="V" tag="VB-P---3P-AA---" g="-" n="P" c="-"
  p="3" m="P">jsou</w>
 <w lemma="věčný" pos="A" tag="AAFP1----1A----" g="F" n="P"
  c="1" p="-" m="-">věčné</w>
...
</s>
```

3 ILLUSTRATIVE ANALYSIS USING THE TXM TOOL

Due to the limited extent of this paper, it is impossible to conduct a more complex analysis, but we shall attempt to illustrate the usefulness of some of the extended options of the TXM tool for corpus analysis of texts. We will move from standard data mining, based on queries for concordances, frequency dictionaries, collocations and other similar phenomena, to multidimensional text analysis.

We randomly selected several texts written by seven authors, particularly:

- Březina – *Hudba pramenů* [Music of the Springs]
- Klíma – *Svět jako vědomí a nic* [World as Consciousness and Nothing]
- Čapek K – *Marsyas* [Marsyas]
- Čapek K – *O umění a kultuře* [On Art and Culture]
- Čapek K – *Kritika slov* [A Critique of Language]
- Čapek J – *Kulhavý poutník* [The Lame Pilgrim]
- Čapek J – *Nejskromnější umění* [The Humblest Art]
- Čapek J – *Co má člověk z umění* [What Man Gets from Art]
- Durych – *Essaye* [Essays]
- Mathesius – *Kulturní aktivismus* [Cultural Activism]
- Šalda – *Boje o zítřek* [Battles for Tomorrow]
- Šalda – *Duše a dílo* [Soul and Work]

The R tool implemented in TXM enables us to use two types of multidimensional analysis: (1) *cluster analysis*, which produces dendrograms expressing similarities or differences between the individual entities compared (text, author, genre, etc.), and (2) *factorial correspondence analysis*.[6] Both these types of quantitative analysis enable comparing the *p-attributes*, i.e. not only lexical items (`word`, `lemma`), but also grammatical categories (part of speech, gender, person, etc.). Apart from lexical analysis, one of the morphological categories showing interesting results using the data sample from CCE is e.g. the grammatical category of number (a tendency for

[6] For more detailed information about both types of analysis, see TXM Manual – ref. [14], pp. 107ff., and reference [15].

usage of singular forms in the case of Šalda, and a very strong tendency for utilization of the plural in the case of Březina). Concerning the parts of speech, an author that differs more significantly from the others is Klíma; the reason for the difference is, however, very specific (see below).

3.1 Ad 1 – Analysis of clusters: Dendrograms

Fig. 1 presents the result of a cluster analysis regarding the main parts of speech for the individual authors.

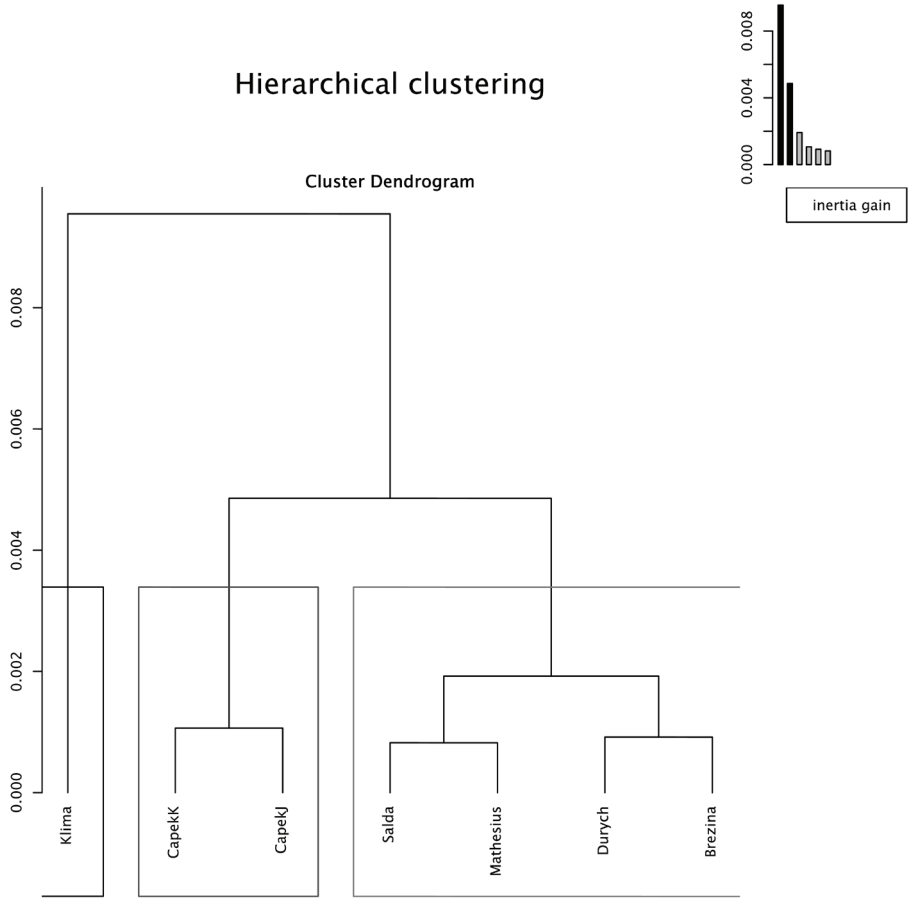

Fig. 1. Dendrogram – the main parameter: part of speech (POS); data: corpus CCE; tool TXM

It is apparent that the biggest difference is that between Klíma and other authors, which is also confirmed by additional three-cluster sub-analyses with the parameters `lemma`, `word`, `pos`, and `tag` (Table 1):

parameter of analysis	cluster 1	cluster 2	cluster 3
`lemma` and `word`	Klíma	Šalda	others
`pos` and `tag`	Klíma	Čapek brothers	others
`n` (grammatical number)	Březina	Mathesius, Šalda, Durych	Klíma, Čapek brothers

Tab. 1. Three-cluster sub-analyses with parameters `lemma`, `word`, `pos`, and `tag`; data: corpus CCE; tool TXM

The reason for the difference in the texts written by Klíma from the rest of the authors is documented in the following correspondence analysis, which also shows the reason for Březina's difference with regard to the grammatical number, which is complemented with a visualization of the specificity score analysis.

3.2 Ad 2 – Factorial correspondence analysis

Using the p-attribute `pos` even for the subsequent correspondence analysis, we can identify a rather specific reason for Klíma's difference: in fact, it is not a POS category, but instead a difference in punctuation (see the tag Z for Klíma):

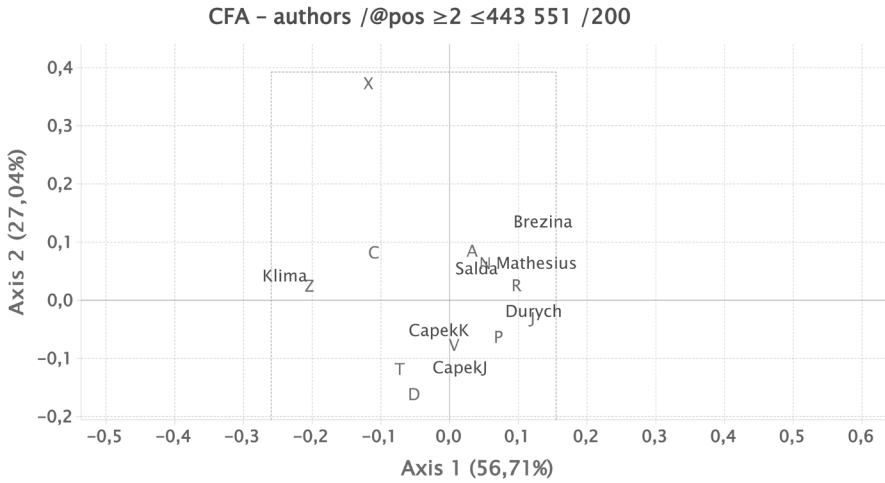

Fig. 2. Factorial analysis – the main parameter: part of speech (`pos`); data: corpus CCE; tool TXM

Klíma's manner of using punctuation marks is highly specific, as illustrated in Fig. 3: m-dashes combined in various ways with a sequence of periods, a semi-colon, or a colon (further combined with a period, a comma, a question mark, or an exclamation mark).


L. Klíma: Svět jako vědomí a nic; I. Všeobecné (1. vyd. 1904); odd. 24

24.

— Nejrozumnější však maxima pro tvory, kteří mohou jednání své vědomě řídit, byla by: „*Dělej cokoli*!“ —: Jest nanejvýš lhostejno, jak jednáš, nejen pro svět, ale i pro tebe! —: jest předem jisto, jak vše dopadne: — — *Každý myšlenkový atom sloučí se průběhem světového roku se* **všemi** *myšlenkovými atomy a* **stane se tak za dobu světového roku tím, čím je svět v každém momentu**; — atom myšlenkový je v témže poměru k světu, jako atom časový k světovému roku. — Dále: *Z výlučné intellektualnosti světa následuje výlučná realnost* **všech** *pomyšlení*, následuje, že není představy bez **realního** důvodu, přání bez **skutečného** objektu; — *že každé pomyšlení jest jen poukazem na skutečnost, každé přání na vyplnění*; — **nic neleží mimo železný kruh vůle v tomto veskrze intellektualním světě**; — nic není irrealního v tomto irrealním světě —. Poněvadž se však každý myšlenkový atom sloučí se všemi ostatními, následuje, **že všechna přání a pomyšlení musí jednou dojít splnění a uskutečnění**! Konečně: Jako positiva a negativa abstracta „svět“ ruší se každým okamžikem, **tak ruší se positiva a negativa každé části každým světovým rokem**: neboť kombinace jednoho atomu se všemi, stejným dílem positivními a negativními, musí být rovněž stejným dílem positivní a negativní; následuje: **svět je vůbec bez hodnoty**... — —: — — Všeho dosáhneš, a vše dosáhne tebe, dělej co dělej! Tvá práce je zbytečna, tvá lenost bez významu, — tak i tak dosáhneš všeho!... *Všechna* tvá nejsmělejší přání se splní, všechny tvé nejhroznější obavy se uskuteční, všechny tvé nejvzdušnější fantasie stanou se realnosti! Povzneseš se k nejvyššímu, klesneš k nejhlubšímu! Nemyslitelné stane se tvou myšlenkou, a nemožné stane se ti skutečností...! Prožiješ všechny metamorfosy, které jsi schopen si představit, a millionkrát více těch, které jsou ti nepředstavitelný!... Ale co budeš mít z tohoto všeho?: — Praničeho! Vším nestaneš se ani šťastnější, ani nešťastnější, ani větší, ani menší... — Všechno tvé úsilné hledání štěstí nerozmnoží ho, všechno tvé odříkání nezmenší ho! Tvé obavy před bolestí jsou nesmyslny: trpíš-li, raduj se, žes si toho zas už kus odbyl; raduješ-li se, věští to pouze bolest, *přesně* tím větší, čím větší tvá radost! Nanejvýš lhostejno jest nejen, co činíš, ale i co se ti přihodí!: jednou musíš si vše odbýt! Tvé úsilí zlepšit tvůj stav je směšné, radost rovněž! každé zlepšení zaplatíš až do posledního haléře zhoršením, štěstí neštěstím, velkost malostí; ale každé tvé špatno promění se v dobro. Nesmíš v *nic doufat*, ale nemusíš se *ničeho obávat*! — Až bude závratný koloběh všeho u konce, jaký bude výsledek...?: *Žádný* ..., — *jako by nic nebylo bývalo*...: *jediná cena všeho* se zatím v koloběhu tom *rozplynula*... A pak začne tento příšerný svět otáčet se znovu.... a po věčnosti věčností... — — Krátce před svou smrtí snil Lichtenberg, že zavítal do vesnické krčmy. Mladý jakýs muž jedl tam polévku, občas vyhazoval ji do výše a lžící opět chytal. Jiní mužové hráli tam v kostky; vedle nich pletla vysoká, hubená žena. Jí ptal se L., možno-li zde co vyhrát? Na odpověď: „nic“ — tázal se pak, možno-li zde co prohrát, a dostal odpověď: „ničeho“. „To považoval jsem za důležitou hru,“ — praví k tomu L. — —: Ano, za důležitou považujeme životní hru, při níž na konec nemůžeme ani ztratit ani získat! Zde je nejvnitřnější *hrozné tajemství* tohoto světa — fantomu: Vše snaží se a plahočí, touží a děsí se, doufá a zoufá, jásá a naříká pro něco, co nejen theoreticky „jest“ *ničím*, ale co i prakticky **v nic** se *paralysuje*...

Fig. 3. Original and specific punctuation of L. Klíma; source: the book *Svět jako vědomí a nic* [The World as Consciousness and Nothing] (1904)

This multidimensional analysis enabled us to detect a highly interesting factor of Klíma's punctuation (worthy of further analysis). It would be appropriate, however, to consider even filtering out this category, which could result in higher precision of the part-of-speech analysis.

A correspondence analysis of the grammatical number also reveals interesting results, where a similar deviation of Otokar Březina from other authors may be observed in the Fig. 4.

A highly useful function that may explain the reason for this obvious difference is the "specificity score" ([14], [16]), which could also be used as one of the alternative approaches to the extraction of prominent text units (thematic expressions,

keywords, etc.).[7] It belongs to the so-called adjusted frequencies which should reflect the actual dispersion or prominence of language expressions or categories in texts, i.e. express their importance rate in the form of hierarchical lists of frequency distribution. TXM even enables a practical option of visualization of these factors, which is helpful when interpreting the results (see Fig. 5).

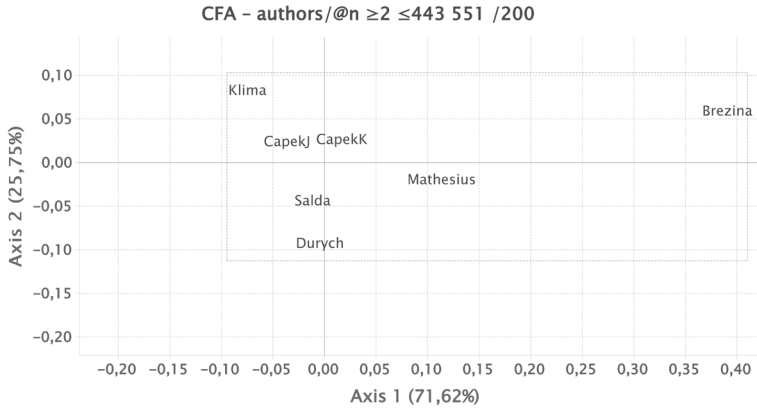

Fig. 4. Factorial analysis – the main parameter: grammatical number (n); data: corpus CCE; tool TXM

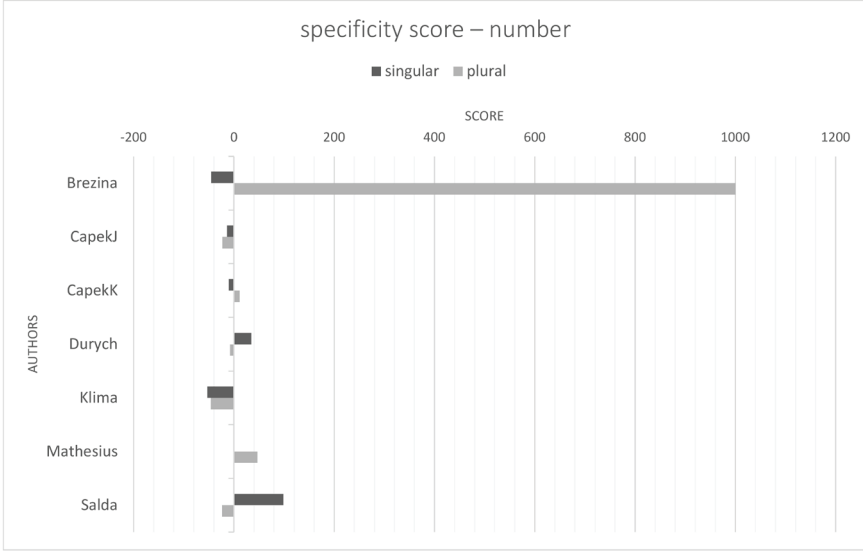

Fig. 5. Specificity score – the main parameter: grammatical number (n); data: corpus CCE; tool TXM

[7] For more information about this quantitative index (including the mathematical formula for its calculation), see TXM Manual [14], pp. 95ff., and reference [16].

The graph clearly indicates a strong, obvious tendency for usage of the plural in the case of Březina (collective plural). We can also observe a very slight tendency for more frequent singular forms in the case of Šalda (subjectivism).

We shall now complement the analysis and the usage of the specificity score with the distribution of the autosemantic parts of speech for the individual authors (we have even added pronouns, as it is an important category for literary texts).

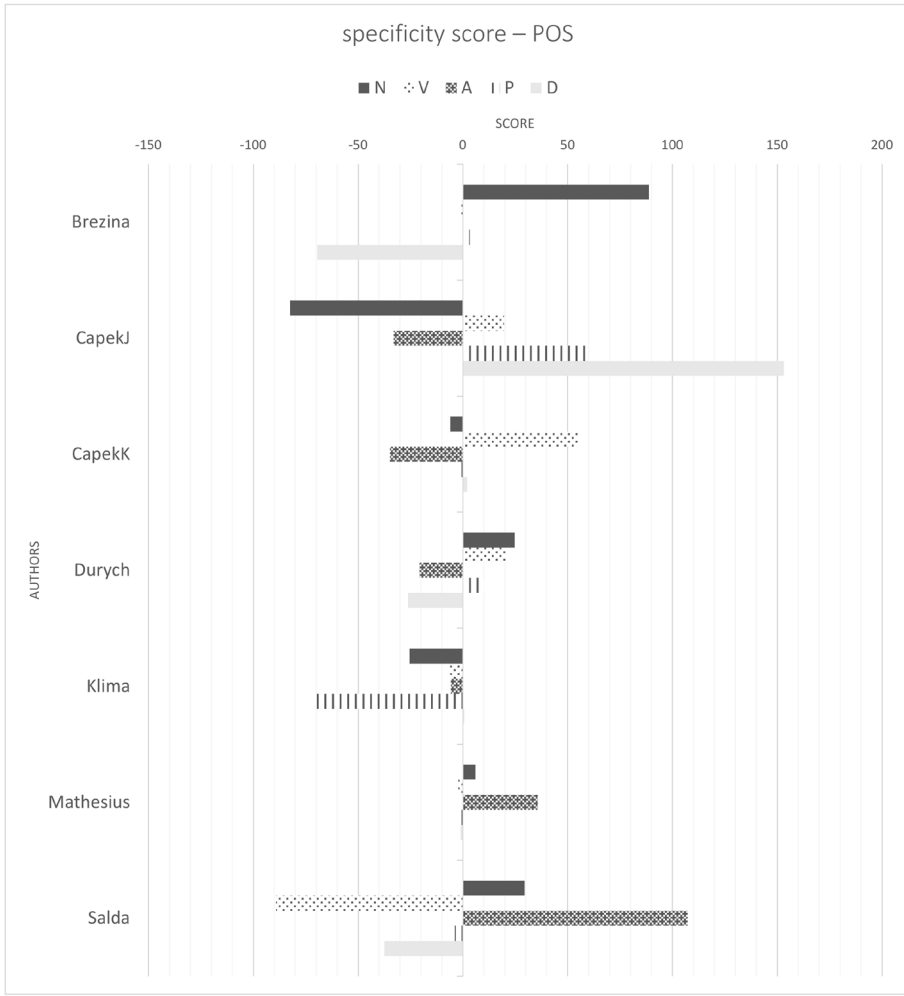

Fig. 6. Specificity score – the main parameter: part of speech (`pos`); data: corpus CCE; tool TXM

Explanatory note: `A` = adjective, `D` = adverb, `N` = noun, `P` = pronoun, `V` = verb.

A visualization of the specificity index reveals the following tendencies in the language of the compared authors:

- The category of nouns is important especially in the case of Březina, to a certain extent also in the case of Durych and Šalda (nominal form of expression, the effort to name substances), in contrast to a rather significant deficit in utilization of nouns in the case of Josef Čapek, compared to the others.
- Adjectives play an important role in the case of Šalda, as well as Mathesius, while a slight deficit is obvious with the Čapek brothers and Durych.
- Pronouns are overused by Josef Čapek, while Klíma suppresses their utilization in his texts.
- Verbs are a dominant part of speech in the case of Karel Čapek, while with Šalda we can see a rather surprising and significant deficit in verbs.
- Adverbs are the most significant part of speech for Josef Čapek, while the opposite tendency may be identified in the case of Březina, and to a certain extent also with Durych and Šalda.

Put simply, we may argue that nouns are the most important and most dominant part of speech in Březina's texts, as with adjectives in the case of Šalda and to a lesser extent also Mathesius. Verbs, a dynamic part of speech, are used to the largest extent by Karel Čapek. In the texts of Josef Čapek, there is a need to focus in greater detail on the prevalence of adverbs, as well as on pronouns. Once again, Klíma is an interesting author: in his texts we can find a deficit in the utilization of nouns and especially pronouns, compared to other authors.

This type of analysis may subsequently serve as background for additional, more traditional, corpus explorations. The findings resulting from this probe enable further analysis to be targeted and focused on more specific phenomena, and especially on those that prove to be relevant or interesting in the texts we are dealing with.

4 CONCLUSION

One of the main aims of the presented project was to establish a linguistically annotated corpus database of Czech literary essays from the turn of the 1900s (we expect that the database will gradually be expanded with new texts and authors). The period from 1890 until the 1930s or 1940s was not chosen randomly: it is a period when the literary essay was formed as a specific, autonomous, and valuable language unit. In addition, the period saw discussions, polemics, and formation of the orthographic form of Czech. This database may therefore serve as a convenient tool for language analyses capturing this development and formation of one language and literary unit.

ACKNOWLEDGEMENTS

The research was supported by the Ministry of Education of the Czech Republic IGA_FF_2020_021 "Czech Studies: Literary and Linguistic Overlaps and Interpretations".

References

[1] Taxová, E. (1985). Experimenty. Český literární esej z přelomu 19. a 20. století. Praha: Melantrich.
[2] Opelík, J. (1986). Lehký harcovník. Antologie českého literárního eseje 2. Léta desátá a dvacátá 20. století. Praha: Melantrich.
[3] Havránek, B. (1932). Úkoly spisovného jazyka a jeho kultura. In Spisovná čeština a jazyková kultura, pages 32–84, Praha.
[4] Jedlička, A. (1989). K jazyku a stylu českých esejistických textů, Slovo a slovesnost, 50, pages 114–125.
[5] Hausenblas, K. (1972). Učební styl v soustavě stylů funkčních, Naše řeč, 55, pages 150–158.
[6] Jelínek, M., and Krčmová, M. (2017). Esejistický styl. In CzechEncy – Nový encyklopedický slovník češtiny. Available at: https://www.czechency.org/slovnik/ESEJISTICKÝ STYL.
[7] Jedlička, A. (1973). K vymezení a charakteristice esejistického stylu, Studia Slavica Pragensia, pages 167–178.
[8] Mistrík, J. (1974). Esejistický štýl, Slovenská reč, 40, pages 321–332.
[9] Digital library Kramerius: Available at: http://kramerius.nkp.cz.
[10] Digital Library MZK: Available at: http://www.digitalniknihovna.cz/mzk.
[11] Textometrie project: TXM (version 0.8.1) [Software]. Available at: http://textometrie.ens-lyon.fr.
[12] The R Project for Statistical Computing: R (version 4.0.4) [Software]. Available at: https://www.r-project.org/.
[13] Morphological Dictionary and Tagger: MorphoDiTa [Software]. Available at: http://lindat.mff.cuni.cz/services/morphodita/.
[14] TXM Reference manual (v0.7). Available at: http://textometrie.ens-lyon.fr/files/documentation/TXMManual 0.7.pdf.
[15] Benzécri, J.-P. et al. (1973). L'analyse des correspondances. Paris: Dunod.
[16] Lafon, P. (1980). Sur la variabilité de la fréquence des formes dans un corpus, Mots, 1, pages 127–165.